

U.S. Serial No.: 10/549,871
Examiner: Mohammad Y. Meah
GAU: 1652
Page 8 of 14

RECEIVED
CENTRAL FAX CENTER

AUG 13 2007

STATUS OF CLAIMS

Claims 1-10, 15-18, 23-24 and 29 are pending, claims 1 and 29 being the sole independent claims. Claims 11-14, 19-22, and 30 have been withdrawn as being drawn to non-elected inventions.

REMARKS

Claims 1-10, 15-18, 23-24 and 29 are pending and rejected. The Applicant has amended Claims 1-2, 4-5, 9, 15, 17, 24 and 29 and added new claims 31 and 32. No new matter has been added by amending the claims. Support for such amendments can be found throughout the specification and specifically on pages 3, 4 and 8 of the specification.

Objection to Claims 1-10, 23-24 and 29 - informalities

The objection to 1-10, 23-24 and 29 is made moot by the amendment to these claims herein.

Rejection Under 35 U.S.C. 112, first paragraph

Claims 1-10, 15-18, 23-24 and 29 were rejected under 35 U.S.C. §112, first paragraph. The Applicant respectfully disagrees.

A. Written Description requirement Rejection

The Office rejected Claims 1-10, 15-18, 23-24 and 29 under 35 U.S.C. 112, first paragraph as allegedly failing to describe in any fashion the "physical and/or chemical properties of the claimed class of biosynthetic intermediates and/or metabolites thereof or class of ergostanol-biosynthetic enzymes including squalene epoxidase..."

In order to allow further prosecution on the merits, the Applicant has amended Claims 1, 24 and 29 to remove the phrase "and/or its biosynthetic intermediates and/or metabolites" and inserting instead specific biosynthetic intermediates and metabolites of ergosta-5,7-dienol. The intermediates and metabolites added to the claims do not constitute new matter because the scope of the amended claims is no broader than (and, thus, is supported by) the

U.S. Serial No.: 10/549,871
Examiner: Mohammad Y. Meah
GAU: 1652
Page 9 of 14

specification as filed. Further support may be found on pages 23, lines 26-29 and page 24, lines 35-37.

With respect to the Office's rejection to "class of ergosterol-biosynthetic enzymes including squalene epoxidase", the Applicant respectfully disagrees.

The Applicant amended the claims to specifically identify the intermediates mevalonate, farnesyl pyrophosphate, geraniol pyrophosphate, squalene epoxide, 4-dimethylcholesta-8,14,24-trienol, 4,4-dimethylzymosterol, squalene, farnesol, geraniol, lanosterol, zymosterone and zymosterol and metabolites campesterol, pregnenolone, 17-OH pregnenolone, progesterone, 17-OH-progesterone, 11-deoxycortisol, hydrocortisone, deoxycorticosterone and/or corticosterone of ergosta-5,7-dienol. These enzymes are well known in the art to have specific biochemical reactivity, which is independent of the source and the detail structure. Furthermore, it is well established law that the written description doctrine (as it relates to biotechnology) requires "a precise definition of, such as by, structure, formula, chemical name, or physical properties, not a mere wish or plan for obtaining the claimed chemical invention." See *Regents of the University of California vs. Eli Lilly & Co.*, 119 F.3d, 1559 (1566) (Fed.Cir. 1997). As stated above, the claims have been amended to claim the intermediates and metabolites of ergosta-5,7-dienol and thereby satisfying the written description requirement.

Based upon these amendments, the Applicant respectfully requests that this rejection be withdrawn as it is made moot.

B. Enablement Rejection

The Office rejected Claims 1-10, 15-18, 23-24 and 29 under 35 U.S.C. 112, first paragraph as allegedly not enabling the process of producing ergosta-5,7 dienol, or any biosynthetic intermediates and/or metabolites with a squalene expoxidase having 30% sequence identity to SEQ ID NO:8. The Applicant respectfully disagrees. However, to allow further prosecution on the merits, the Applicant has deleted reference to "30% sequence identity to SEQ ID NO: 8". The Applicant reserves the right to prosecute this invention in a later filed application.

U.S. Serial No.: 10/549,871
Examiner: Mohammad Y. Meah
GAU: 1652
Page 10 of 14

The Office further indicated that the scope of the claims is not commensurate with the disclosure with regard to the "extremely large number of processes of producing the claimed compounds. The Applicant respectfully disagrees.

In making the rejection, the Examiner has specifically contended that

"Since the amino acid sequence of a protein determines its structural and functional properties, predictability of which changes can be tolerated in a protein's amino acid sequence and obtain the desired activity requires a knowledge of and guidance with regard to which amino acids in the protein's sequence, if any are tolerant of modification and which are conserved (i.e., expectedly intolerant to modification), and detailed knowledge of the ways in which the proteins' structure relates to its function." (Office Action page 7)

The first paragraph of 35 U.S.C. § 112 requires that the specification describe how to make and use the claimed subject matter. That requirement has been met in the present application. In particular, the specification describes how to make ergosta-5,7-dienol, its intermediates and metabolites. Throughout the specification there is mention of several different techniques in which to obtain the desired activity of $\Delta 22$ -desaturase, HM-CoA-reductase, squalene epoxidase activities. Specifically on pages 5-6 of the specification outline the various ways in which to reduce the expression of the nucleic acid encoding $\Delta 22$ -desaturase as follows:

- a) introducing nucleic acid sequences which can be transcribed into an antisense nucleic acid sequence which is capable of inhibiting the $\Delta 22$ -desaturase activity, for example by inhibiting the expression of endogenous $\Delta 22$ -desaturase activity,
- b) overexpressing homologous $\Delta 22$ -desaturase nucleic acid sequences, which lead to cosuppression,
- c) introducing nonsense mutations into the endogen by introducing RNA/DNA oligonucleotides into the organism,
- d) introducing specific DNA-binding factors, for example factors of the zinc finger transcription factor type, which bring about a reduced gene expression, or
- e) generating knock-out mutants, for example with the aid of T-DNA mutagenesis or homologous recombination.

U.S. Serial No.: 10/549,871
Examiner: Mohammad Y. Meah
GAU: 1652
Page 11 of 14

In a preferred embodiment of the method according to the invention, the gene expression of the nucleic acids encoding a $\Delta 22$ -desaturase is reduced by generating knockout mutants, especially preferably by homologous recombination.

Likewise, on page pages 11-12 of the specification HMG-CoA-reductase, squalene epoxidase activities, to name a few, identifies several methods in which to accomplish the desired activity in order to produce ergosta-5,7-dienol. For example:

- 1) modifying the promoter DNA sequences,
- 2) increasing or modifying gene expression by applying exogenous stimuli, for example:
 - a) foreign substances, and
 - b) regulator protein (e.g., chimeric protein) that may be for genes encoding.

Thus, one skilled in the art would know how to produce such activities having at least at least 5%, more preferably to at least 20%, more preferably to at least 50%, more preferably to at least 100%, even more preferably to at least 300%, especially preferably to at least 500%, in particular to at least 600% of the enzyme activity as compared to the wild type organism. Therefore, regardless of the precise functional characteristics of the enzymes, one can still produce the enzymatic activities using the disclosure provided by the present specification.

Furthermore, skilled artisans in molecular biology have determined that 30% identity is a reliable threshold for establishing evolutionary homology between two sequences aligned over at least 150 residues. See, *Brenner et al.*, "Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships," *Proc. Natl. Acad. Sci. USA*. (1998) 95:6073-6078). For the Office's convenience, the Applicant has provided herewith a copy of the Brenner et al. reference.

Claims 9 and 17 recites, among others, a method for producing ergosta-5,7-dienol and its intermediates and metabolites in an organism with an increased HMG-CoA reductase and squalene epoxidase activities wherein the nucleic acids introduced into the organism have SEQ ID NOs: 4 and 8, respectively and a sequence that is at least 30% homologous to that of the disclosed SEQ ID NOs: 4 and 8, respectively. The Applicant has amended Claims 9 and

U.S. Serial No.: 10/549,871
Examiner: Mohammad Y. Meah
GAU: 1652
Page 12 of 14

17 and added new Claims 31 and 32 to fix the grammar and to more clearly identify the nucleic acids introduced into the organisms.

As set forth in *In re Marzocchi*, 169 USPQ 367, 369 (CCPA 1971):

The first paragraph of § 112 requires nothing more than objective enablement. How such a teaching is set forth, either by the use of illustrative examples or by broad terminology, is of no importance.

As a matter of Patent Office practice, then, a specification disclosure which contains a teaching of the manner and process of making and using the invention in terms which correspond in scope to those used in describing and defining the subject matter sought to be patented must be taken as in compliance with the enabling requirement of the first paragraph of § 112 unless there is reason to doubt the objective truth of the statements contained therein which must be relied on for enabling support.

Contrary to the standard set forth in *Marzocchi*, the Office has failed to provide any reasons why one would doubt that the guidance provided by the present specification would enable one to make and use the cited enzymes. Based upon these arguments, the Applicant respectfully submits that the Office has not put for a *prima facie* case for non-enablement and request that this rejection be withdrawn.

Rejection of Claims 1-8, 15-16, 23-24 and 29 under 35 U.S.C. 102(b)

Claims 1-8, 15-16, 23-24 and 29 have been rejected under 35 U.S.C. 102(b) as allegedly being anticipated by Saunders et al. (EP 0486290), hereinafter "Saunders".

As the Office indicated, Saunders discloses "ergosta-5,7-dienol using various mutant *S. cerevisiae* strains, with reduce expression of desaturase, increase expression of squalene epoxidase, that are further transformed with a gene encoding HMG-CoA-reductase activity". The Applicant respectfully disagrees for the following reasons.

Saunders discusses a method of "increasing the accumulation of squalene and specific sterols in yeast comprising increasing the expression level of a structural gene encoding a polypeptide having HMG-CoA reductase activity". See Saunders, Technical Invention on page 2. Saunders does not teach the reduction of desaturase (*erg5*), increase HMG-CoA

U.S. Serial No.: 10/549,871
Examiner: Mohammad Y. Meah
GAU: 1652
Page 13 of 14

reductase activity, while also increasing squalene epoxidase activity (nor does Saunders teach the activities of lanosterol C 14-methylase or squalene synthetase activities). In addition, the International Preliminary Examination Report ("IPER") for the priority PCT application differentiates the present application from that of Saunders,

"[Saunders] is considered to be the prior art closest to the subject matter of claim 1 and the subject matter of claim 1 therefore differs from D1 in that the method is implemented using organisms which, by comparison with the present claims, *additionally display an increased activity of a further enzyme*, which enzyme can be selected from the group comprising lanosterol-C14-demethylase (erg11), squalene [e]poxidase (erg1) and squalene synthetase (erg9)."

See 4.2 of IPER, emphasis added.

Therefore, Saunders does not teach every element of the claimed invention. Based upon this difference, Saunders does not anticipate the present application. The Applicant respectfully requests that this rejection be withdrawn.

Rejection of Claims 1-8, 15-16, 23-24 and 29 U.S.C. 102(b)

Claims 1-8, 15-16, 23-24 and 29 have been rejected under 35 U.S.C. 102(b) as allegedly being anticipated by Weber et al. (WO-99/16886), hereinafter "Weber".

As the Office indicated, Weber discloses "the method of making ergosta-5,7-dienol using various mutant *S. cerevisiae* strain with reduce expression of desaturase and increased expression of squalene epoxidase and HMG-CoA-reductase". The Applicant respectfully disagrees for the following reasons.

Weber discusses a method of "producing ergosterol and intermediate products thereof by means of recombinant yeast and plasmids for transforming yeasts". See Saunders, Abstract on page 1. The Office indicated that Weber teaches a reduction in expression of desaturase. After review of the Weber reference, it does not appear to disclose a reduced desaturase activity, which is one of the elements of the presently claimed invention. It is respectfully requested that the Office indicate the specific location of such statement. Based

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6073–6078, May 1998
Biochemistry

Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships

STEVEN E. BRENNER^{*†‡}, CYRUS CHOTHIA^{*}, AND TIM J. P. HUBBARD[§]

^{*}MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom; and [§]Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambs CB10 1SA, United Kingdom

Communicated by David R. Davies, National Institute of Diabetes, Bethesda, MD, March 16, 1998 (received for review November 12, 1997)

ABSTRACT Pairwise sequence comparison methods have been assessed using proteins whose relationships are known reliably from their structures and functions, as described in the SCOP database [Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia C. (1995) *J. Mol. Biol.* 247, 536–540]. The evaluation tested the programs BLAST [Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403–410], WU-BLAST2 [Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* 266, 460–480], FASTA [Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444–2448], and SSEARCH [Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* 147, 195–197] and their scoring schemes. The error rate of all algorithms is greatly reduced by using statistical scores to evaluate matches rather than percentage identity or raw scores. The E-value statistical scores of SSEARCH and FASTA are reliable: the number of false positives found in our tests agrees well with the scores reported. However, the P-values reported by BLAST and WU-BLAST2 exaggerate significance by orders of magnitude. SSEARCH, FASTA $k_{\text{up}} = 1$, and WU-BLAST2 perform best, and they are capable of detecting almost all relationships between proteins whose sequence identities are >30%. For more distantly related proteins, they do much less well; only one-half of the relationships between proteins with 20–30% identity are found. Because many homologs have low sequence similarity, most distant relationships cannot be detected by any pairwise comparison method; however, those which are identified may be used with confidence.

Sequence database searching plays a role in virtually every branch of molecular biology and is crucial for interpreting the sequences issuing forth from genome projects. Given the method's central role, it is surprising that overall and relative capabilities of different procedures are largely unknown. It is difficult to verify algorithms on sample data because this requires large data sets of proteins whose evolutionary relationships are known unambiguously and independently of the methods being evaluated. However, nearly all known homologs have been identified by sequence analysis (the method to be tested). Also, it is generally very difficult to know, in the absence of structural data, whether two proteins that lack clear sequence similarity are unrelated. This has meant that although previous evaluations have helped improve sequence comparison, they have suffered from insufficient, imperfectly characterized, or artificial test data. Assessment also has been problematic because high quality database sequence searching attempts to have both sensitivity (detection of homologs) and specificity (rejection of unrelated proteins); however, these complementary goals are linked such that increasing one causes the other to be reduced.

Sequence comparison methodologies have evolved rapidly, so no previously published tests have evaluated modern versions of programs commonly used. For example, parameters in BLAST (1) have changed, and WU-BLAST2 (2)—which produces gapped alignments—has become available. The latest version of FASTA (3) previously tested was 1.6, but the current release (version 3.0) provides fundamentally different results in the form of statistical scoring.

The previous reports also have left gaps in our knowledge. For example, there has been no published assessment of thresholds for scoring schemes more sophisticated than percentage identity. Thus, the widely discussed statistical scoring measures have never actually been evaluated on large databases of real proteins. Moreover, the different scoring schemes commonly in use have not been compared.

Beyond these issues, there is a more fundamental question: in an absolute sense, how well does pairwise sequence comparison work? That is, what fraction of homologous proteins can be detected using modern database searching methods?

In this work, we attempt to answer these questions and to overcome both of the fundamental difficulties that have hindered assessment of sequence comparison methodologies. First, we use the set of distant evolutionary relationships in the SCOP: Structural Classification of Proteins database (4), which is derived from structural and functional characteristics (5). The SCOP database provides a uniquely reliable set of homologs, which are known independently of sequence comparison. Second, we use an assessment method that jointly measures both sensitivity and specificity. This method allows straightforward comparison of different sequence searching procedures. Further, it can be used to aid interpretation of real database searches and thus provide optimal and reliable results.

Previous Assessments of Sequence Comparison. Several previous studies have examined the relative performance of different sequence comparison methods. The most encompassing analyses have been by Pearson (6, 7), who compared the three most commonly used programs. Of these, the Smith-Waterman algorithm (8) implemented in SSEARCH (3) is the oldest and slowest but the most rigorous. Modern heuristics have provided BLAST (1) the speed and convenience to make it the most popular program. Intermediate between these two is FASTA (3), which may be run in two modes offering either greater speed ($k_{\text{up}} = 2$) or greater effectiveness ($k_{\text{up}} = 1$). Pearson also considered different parameters for each of these programs.

To test the methods, Pearson selected two representative proteins from each of 67 protein superfamilies defined by the PIR database (9). Each was used as a query to search the database, and the matched proteins were marked as being homologous or unrelated according to their membership of PIR.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/956073-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

Abbreviation: EPQ, errors per query.

[†]Present address: Department of Structural Biology, Stanford University, Fairchild Building D-109, Stanford, CA 94305-5126

[‡]To whom reprints requests should be addressed. e-mail: brenner@hyper.stanford.edu.

superfamilies. Pearson found that modern matrices and "ln-scaling" of raw scores improve results considerably. He also reported that the rigorous Smith-Waterman algorithm worked slightly better than PASTA, which was in turn more effective than BLAST.

Very large scale analyses of matrices have been performed (10), and Henikoff and Henikoff (11) also evaluated the effectiveness of BLAST and FASTA. Their test with BLAST considered the ability to detect homologs above a predetermined score but had no penalty for methods which also reported large numbers of spurious matches. The Henikoffs searched the SWISS-PROT database (12) and used PROSITE (13) to define homologous families. Their results showed that the BLOSUM62 matrix (14) performed markedly better than the extrapolated PAM-series matrices (15), which previously had been popular.

A crucial aspect of any assessment is the data that are used to test the ability of the program to find homologs. But in Pearson's and the Henikoffs' evaluations of sequence comparison, the correct results were effectively unknown. This is because the superfamilies in PIR and PROSITE are principally created by using the same sequence comparison methods which are being evaluated. Interdependency of data and methods creates a "chicken and egg" problem, and means for example, that new methods would be penalized for correctly identifying homologs missed by older programs. For instance, immunoglobulin variable and constant domains are clearly homologous, but PIR places them in different superfamilies. The problem is widespread: each superfamily in PIR 48.00 with a structural homolog is itself homologous to an average of 1.6 other PIR superfamilies (16).

To surmount these sorts of difficulties, Sander and Schneider (17) used protein structures to evaluate sequence comparison. Rather than comparing different sequence comparison algorithms, their work focused on determining a length-dependent threshold of percentage identity, above which all proteins would be of similar structure. A result of this analysis was the HSSP equation; it states that proteins with 25% identity over 80 residues will have similar structures, whereas shorter alignments require higher identity. (Other studies also have used structures (18-20), but these focused on a small number of model proteins and were principally oriented toward evaluating alignment accuracy rather than homology detection.)

A general solution to the problem of scoring comes from statistical measures (i.e., E-values and P-values) based on the extreme value distribution (21). Extreme value scoring was implemented analytically in the BLAST program using the Karlin and Altschul statistics (22, 23) and empirical approaches have been recently added to FASTA and SSEARCH. In addition to being heralded as a reliable means of recognizing significantly similar proteins (24, 25), the mathematical tractability of statistical scores "is a crucial feature of the BLAST algorithm" (1). The validity of this scoring procedure has been tested analytically and empirically (see ref. 2 and references in ref. 24). However, all large empirical tests used random sequences that may lack the subtle structure found within biological sequences (26, 27) and obviously do not contain any real homologs. Thus, although many researchers have suggested that statistical scores be used to rank matches (24, 25, 28), there have been no large rigorous experiments on biological data to determine the degree to which such rankings are superior.

A Database for Testing Homology Detection. Since the discovery that the structures of hemoglobin and myoglobin are very similar though their sequences are not (29), it has been apparent that comparing structures is a more powerful (if less convenient) way to recognize distant evolutionary relationships than comparing sequences. If two proteins show a high degree of similarity in their structural details and function, it

is very probable that they have an evolutionary relationship though their sequence similarity may be low.

The recent growth of protein structure information combined with the comprehensive evolutionary classification in the SCOP database (4, 5) have allowed us to overcome previous limitations. With these data, we can evaluate the performance of sequence comparison methods on real protein sequences whose relationships are known confidently. The SCOP database uses structural information to recognize distant homologs, the large majority of which can be determined unambiguously. These superfamilies, such as the globins or the immunoglobulins, would be recognized as related by the vast majority of the biological community despite the lack of high sequence similarity.

From SCOP, we extracted the sequences of domains of proteins in the Protein Data Bank (PDB) (30) and created two databases. One (PDB90D-B) has domains, which were all <90% identical to any other, whereas (PDB40D-B) had those <40% identical. The databases were created by first sorting all protein domains in SCOP by their quality and making a list. The highest quality domain was selected for inclusion in the database and removed from the list. Also removed from the list (and discarded) were all other domains above the threshold level of identity to the selected domain. This process was repeated until the list was empty. The PDB40D-B database contains 1,323 domains, which have 9,044 ordered pairs of distant relationships, or ~0.5% of the total 1,749,006 ordered pairs. In PDB90D-B, the 2,079 domains have 53,988 relationships, representing 1.2% of all pairs. Low complexity regions of sequence can achieve spurious high scores, so these were masked in both databases by processing with the SEG program (27) using recommended parameters: 12 1.8 2.0. The databases used in this paper are available from <http://sss.stanford.edu/sss/>, and databases derived from the current version of SCOP may be found at <http://scop.mrc-lmb.cam.ac.uk/scop/>.

Analyses from both databases were generally consistent, but PDB40D-B focuses on distantly related proteins and reduces the heavy overrepresentation in the PDB of a small number of families (31, 32), whereas PDB90D-B (with more sequences) improves evaluations of statistics. Except where noted otherwise, the distant homolog results here are from PDB40D-B. Although the precise numbers reported here are specific to the structural domain databases used, we expect the trends to be general.

Assessment Data and Procedure. Our assessment of sequence comparison may be divided into four different major categories of tests. First, using just a single sequence comparison algorithm at a time, we evaluated the effectiveness of different scoring schemes. Second, we assessed the reliability of scoring procedures, including an evaluation of the validity of statistical scoring. Third, we compared sequence comparison algorithms (using the optimal scoring scheme) to determine their relative performance. Fourth, we examined the distribution of homologs and considered the power of pairwise sequence comparison to recognize them. All of the analyses used the databases of structurally identified homologs and a new assessment criterion.

The analyses tested BLAST (1), version 1.4.9MP, and WU-BLAST2 (2), version 2.0a13MP. Also assessed was the FASTA package, version 3.0t76 (3), which provided PASTA and the SSEARCH implementation of Smith-Waterman (8). For SSEARCH and FASTA, we used BLOSUM45 with gap penalties -12/-1 (7, 16). The default parameters and matrix (BLOSUM62) were used for BLAST and WU-BLAST2.

The "Coverage Vs. Error" Plot. To test a particular protocol (comprising a program and scoring scheme), each sequence from the database was used as a query to search the database. This yielded ordered pairs of query and target sequences with associated scores, which were sorted, on the basis of their scores, from best to worst. The ideal method would have

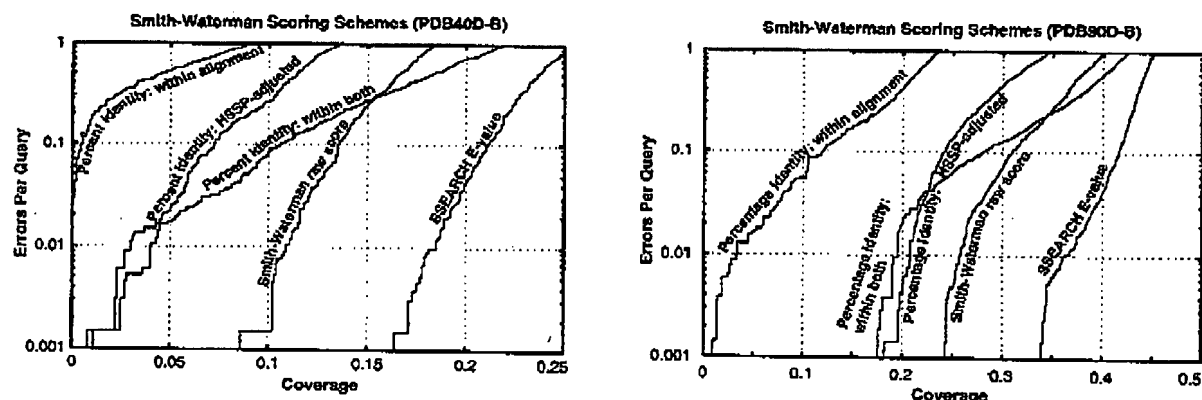


FIG. 1. Coverage vs. error plots of different scoring schemes for SSEARCH Smith-Waterman. (A) Analysis of PDB40D-B database. (B) Analysis of PDB90D-B database. All of the proteins in the database were compared with each other using the SSEARCH program. The results of this single set of comparisons were considered using five different scoring schemes and assessed. The graphs show the coverage and errors per query (EPQ) for statistical scores, raw scores, and three measures using percentage identity. In the coverage vs. error plot, the x axis indicates the fraction of all homologs in the database (known from structure) which have been detected. Precisely, it is the number of detected pairs of proteins with the same fold divided by the total number of pairs from a common superfamily. PDB40D-B contains a total of 9,044 homologs, so a score of 10% indicates identification of 904 relationships. The y axis reports the number of EPQ. Because there are 1,323 queries made in the PDB40D-B all-vs.-all comparison, 13 errors corresponds to 0.01, or 1% EPQ. The y axis is presented on a log scale to show results over the widely varying degrees of accuracy which may be desired. The scores that correspond to the levels of EPQ and coverage are shown in Fig. 4 and Table 1. The graph demonstrates the trade-off between sensitivity and selectivity. As more homologs are found (moving to the right), more errors are made (moving up). The ideal method would be in the lower right corner of the graph, which corresponds to identifying many evolutionary relationships without selecting unrelated proteins. Three measures of percentage identity are plotted. Percentage identity within alignment is the degree of identity within the aligned region of the proteins, without consideration of the alignment length. Percentage identity within both is the number of identical residues in the aligned region as a percentage of the average length of the query and target proteins. The HSSP equation (17) is $H = 290.15I^{-0.562}$ where I is length for $10 < I < 80$; $H > 100$ for $I < 10$; $H = 24.7$ for $I > 80$. The percentage identity HSSP-adjusted score is the percent identity within the alignment minus H . Smith-Waterman raw scores and E-values were taken directly from the sequence comparison program.

perfect separation, with all of the homologs at the top of the list and unrelated proteins below. In practice, perfect separation is impossible to achieve so instead one is interested in drawing a threshold above which there are the largest number of related pairs of sequences consistent with an acceptable error rate.

Our procedure involved measuring the coverage and error for every threshold. Coverage was defined as the fraction of structurally determined homologs that have scores above the selected threshold; this reflects the sensitivity of a method. Errors per query (EPQ), an indicator of selectivity, is the number of nonhomologous pairs above the threshold divided by the number of queries. Graphs of these data, called coverage vs. error plots, were devised to understand how

protocols compare at different levels of accuracy. These graphs share effectively all of the beneficial features of Receiver Operating Characteristic (ROC) plots (33, 34) but better represent the high degrees of accuracy required in sequence comparison and the huge background of nonhomologs.

This assessment procedure is directly relevant to practical sequence database searching, for it provides precisely the information necessary to perform a reliable sequence database search. The EPQ measure places a premium on score consistency; that is, it requires scores to be comparable for different queries. Consistency is an aspect which has been largely

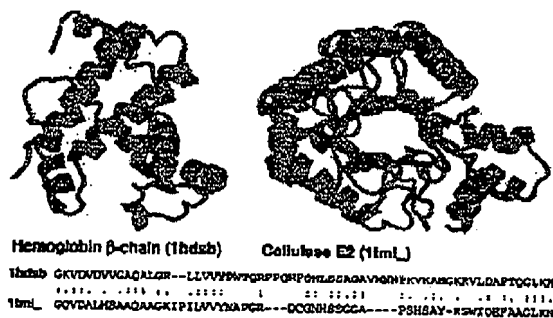


FIG. 2. Unrelated proteins with high percentage identity. Hemoglobin β -chain (PDB code 1hds chain b, ref. 38, Left) and cellulase E2 (PDB code 1tml, ref. 39, Right) have 39% identity over 64 residues, a level which is often believed to be indicative of homology. Despite this high degree of identity, their structures strongly suggest that these proteins are not related. Appropriately, neither the raw alignment score of 85 nor the E-value of 1.3 is significant. Proteins rendered by RASMOI, (40).

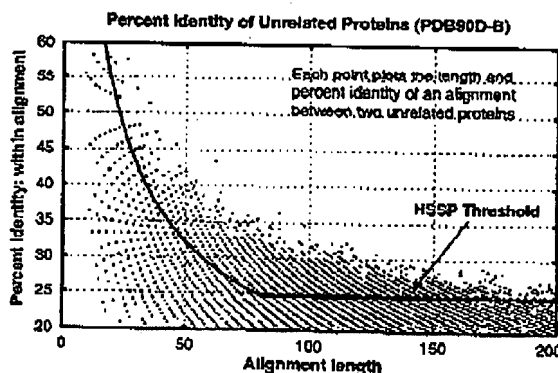


FIG. 3. Length and percentage identity of alignments of unrelated proteins in PDB90D-B. Each pair of nonhomologous proteins found with SSEARCH is plotted as a point whose position indicates the length and the percentage identity within the alignment. Because alignment length and percentage identity are quantized, many pairs of proteins may have exactly the same alignment length and percentage identity. The line shows the HSSP threshold (though it is intended to be applied with a different matrix and parameters).

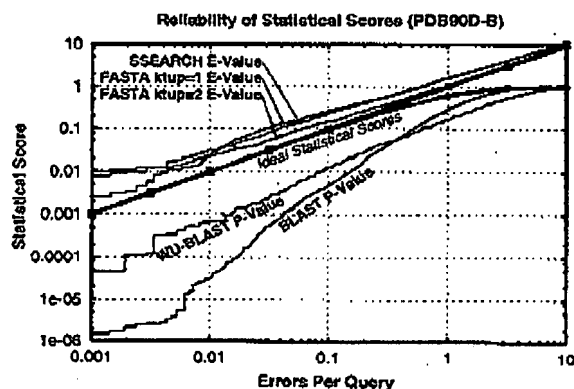


FIG. 4. Reliability of statistical scores in PDB90D-B: Each line shows the relationship between reported statistical score and actual error rate for a different program. E-values are reported for SSEARCH and FASTA, whereas P-values are shown for BLAST and WU-BLAST2. If the scoring were perfect, then the number of errors per query and the E-values would be the same, as indicated by the upper bold line. (P-values should be the same as EPQ for small numbers, and diverges at higher values, as indicated by the lower bold line.) E-values from SSEARCH and FASTA are shown to have good agreement with EPQ but underestimate the significance slightly. BLAST and WU-BLAST2 are overconfident, with the degree of exaggeration dependent upon the score. The results for PDB40D-B were similar to those for PDB90D-B despite the difference in number of homologs detected. This graph could be used to roughly calibrate the reliability of a given statistical score.

ignored in previous tests but is essential for the straightforward or automatic interpretation of sequence comparison results. Further, it provides a clear indication of the confidence that should be ascribed to each match. Indeed, the EPQ measure should approximate the expectation value reported by database searching programs, if the programs' estimates are accurate.

The Performance of Scoring Schemes. All of the programs tested could provide three fundamental types of scores. The first score is the percentage identity, which may be computed in several ways based on either the length of the alignment or the lengths of the sequences. The second is a "raw" or "Smith-Waterman" score, which is the measure optimized by the Smith-Waterman algorithm and is computed by summing the substitution matrix scores for each position in the alignment and subtracting gap penalties. In BLAST, a measure

related to this score is scaled into bits. Third is a statistical score based on the extreme value distribution. These results are summarized in Fig. 1.

Sequence Identity. Though it has been long established that percentage identity is a poor measure (35), there is a common rule-of-thumb stating that 30% identity signifies homology. Moreover, publications have indicated that 25% identity can be used as a threshold (17, 36). We find that these thresholds, originally derived years ago, are not supported by present results. As databases have grown, so have the possibilities for chance alignments with high identity; thus, the reported cutoffs lead to frequent errors. Fig. 2 shows one of the many pairs of proteins with very different structures that nonetheless have high levels of identity over considerable aligned regions. Despite the high identity, the raw and the statistical scores for such incorrect matches are typically not significant. The principal reasons percentage identity does so poorly seem to be that it ignores information about gaps and about the conservative or radical nature of residue substitutions.

From the PDB90D-B analysis in Fig. 3, we learn that 30% identity is a reliable threshold for this database only for sequence alignments of at least 150 residues. Because one unrelated pair of proteins has 43.5% identity over 62 residues, it is probably necessary for alignments to be at least 70 residues in length before 40% is a reasonable threshold, for a database of this particular size and composition.

At a given reliability, scores based on percentage identity detect just a fraction of the distant homologs found by statistical scoring. If one measures the percentage identity in the aligned regions without consideration of alignment length, then a negligible number of distant homologs are detected. Use of the HSP equation improves the value of percentage identity, but even this measure can find only 4% of all known homologs at 1% EPQ. In short, percentage identity discards most of the information measured in a sequence comparison.

Raw Scores. Smith-Waterman raw scores perform better than percentage identity (Fig. 1), but ln-scaling (7) provided no notable benefit in our analysis. It is necessary to be very precise when using either raw or bit scores because a 20% change in cutoff score could yield a tenfold difference in EPQ. However, it is difficult to choose appropriate thresholds because the reliability of a bit score depends on the lengths of the proteins matched and the size of the database. Raw score thresholds also are affected by matrix and gap parameters.

Statistical Scores. Statistical scores were introduced partly to overcome the problems that arise from raw scores. This scoring scheme provides the best discrimination between homologous proteins and those which are unrelated. Most

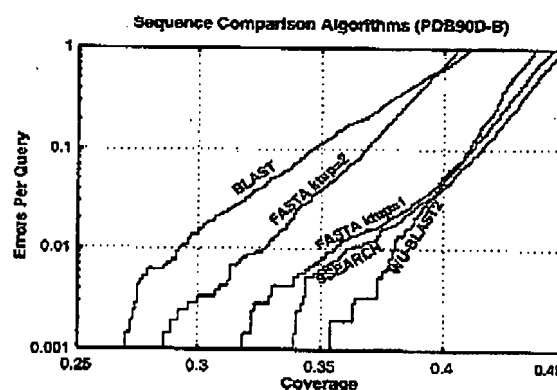
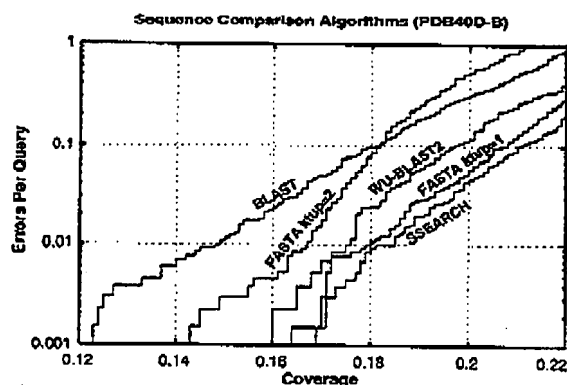


FIG. 5. Coverage vs. error plots of different sequence comparison methods: Five different sequence comparison methods are evaluated, each using statistical scores (E- or P-values). (A) PDB40D-B database. In this analysis, the best method is the slow SSEARCH, which finds 18% of relationships at 1% EPQ. FASTA kmp-1 and WU-BLAST2 are almost as good. (B) PDB90D-B database. The quick WU-BLAST2 program provides the best coverage at 1% EPQ on this database, although at higher levels of error it becomes slightly worse than FASTA kmp-1 and SSEARCH.

likely, its power can be attributed to its incorporation of more information than any other measure; it takes account of the full substitution and gap data (like raw scores) but also has details about the sequence lengths and composition and is scaled appropriately.

We find that statistical scores are not only powerful, but also easy to interpret. SSEARCH and FASTA show close agreement between statistical scores and actual number of errors per query (Fig. 4). The expectation value score gives a good, slightly conservative estimate of the chances of the two sequences being found at random in a given query. Thus, an E-value of 0.01 indicates that roughly one pair of nonhomologs of this similarity should be found in every 100 different queries. Neither raw scores nor percentage identity can be interpreted in this way, and these results validate the suitability of the extreme value distribution for describing the scores from a database search.

The P-values from BLAST also should be directly interpretable but were found to overstate significance by more than two orders of magnitude for 1% EPQ for this database. Nonetheless, these results strongly suggest that the analytic theory is fundamentally appropriate. WU-BLAST2 scores were more reliable than those from BLAST, but also exaggerate expected confidence by more than an order of magnitude at 1% EPQ.

Overall Detection of Homologs and Comparison of Algorithms. The results in Fig. 5A and Table 1 show that pairwise sequence comparison is capable of identifying only a small fraction of the homologous pairs of sequences in PDB40D-B. Even SSEARCH with E-values, the best protocol tested, could find only 18% of all relationships at a 1% EPQ. BLAST, which identifies 15%, was the worst performer, whereas FASTA $k_{\text{up}} = 1$ is nearly as effective as SSEARCH. FASTA $k_{\text{up}} = 2$ and WU-BLAST2 are intermediate in their ability to detect homologs. Comparison of different algorithms indicates that those capable of identifying more homologs are generally slower. SSEARCH is 25 times slower than BLAST and 6.5 times slower than FASTA $k_{\text{up}} = 1$. WU-BLAST2 is slightly faster than FASTA $k_{\text{up}} = 2$, but the latter has more interpretable scores.

In PDB90D-B, where there are many close relationships, the best method can identify only 38% of structurally known homologs (Fig. 5B). The method which finds that many relationships is WU-BLAST2. Consequently, we infer that the differences between FASTA $k_{\text{up}} = 1$, SSEARCH, and WU-BLAST2 programs are unlikely to be significant when compared with variation in database composition and scoring reliability.

Fig. 6 helps to explain why most distant homologs cannot be found by sequence comparison: a great many such relationships have no more sequence identity than would be expected by chance. SSEARCH with E-values can recognize >90% of the homologous pairs with 30–40% identity. In this region, there are 30 pairs of homologous proteins that do not have significant E-values, but 26 of these involve sequences with <50 residues. Of sequences having 25–30% identity, 75% are identified by SSEARCH E-values. However, although the number of homologs grows at lower levels of identity, the detection falls off sharply: only 40% of homologs with 20–25% identity

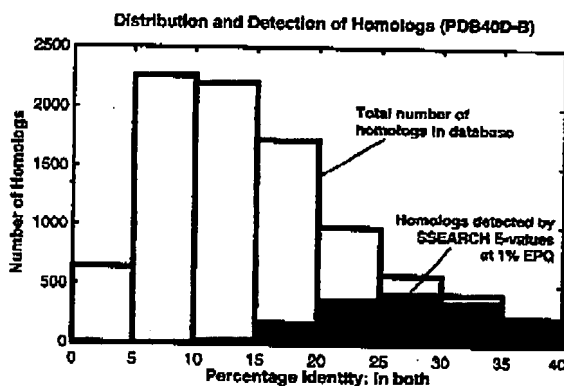


Fig. 6. Distribution and detection of homologs in PDB40D-B. Bars show the distribution of homologous pairs PDB40D-B according to their identity (using the measure of identity in both). Filled regions indicate the number of these pairs found by the best database searching method (SSEARCH with E-values) at 1% EPQ. The PDB40D-B database contains proteins with <40% identity, and as shown on this graph, most structurally identified homologs in the database have diverged extremely far in sequence and have <20% identity. Note that the alignments may be inaccurate, especially at low levels of identity. Filled regions show that SSEARCH can identify most relationships that have 25% or more identity, but its detection wanes sharply below 25%. Consequently, the great sequence divergence of most structurally identified evolutionary relationships effectively defeats the ability of pairwise sequence comparison to detect them.

are detected and only 10% of those with 15–20% can be found. These results show that statistical scores can find related proteins whose identity is remarkably low; however, the power of the method is restricted by the great divergence of many protein sequences.

After completion of this work, a new version of pairwise BLAST was released: BLASTGP (37). It supports gapped alignments, like WU-BLAST2, and dispenses with sum statistics. Our initial tests on BLASTGP using default parameters show that its E-values are reliable and that its overall detection of homologs was substantially better than that of ungapped BLAST, but not quite equal to that of WU-BLAST2.

CONCLUSION

The general consensus amongst experts (see refs. 7, 24, 25, 27 and references therein) suggests that the most effective sequence searches are made by (i) using a large current database in which the protein sequences have been complexity masked and (ii) using statistical scores to interpret the results. Our experiments fully support this view.

Our results also suggest two further points. First, the E-values reported by FASTA and SSEARCH give fairly accurate estimates of the significance of each match, but the P-values provided by BLAST and WU-BLAST2 underestimate the true

Table 1. Summary of sequence comparison methods with PDB40D-B

Method	Relative Time*	1% EPQ Cutoff	Coverage at 1% EPQ
SSEARCH % identity: within alignment	25.5	>70%	<0.1
SSEARCH % identity: within both	25.5	34%	3.0
SSEARCH % identity: HSP-scaled	25.5	35% (HSP + 9.8)	4.0
SSEARCH Smith-Waterman raw scores	25.5	142	10.5
SSEARCH E-values	25.5	0.03	18.4
FASTA $k_{\text{up}} = 1$ E-values	3.9	0.03	17.9
FASTA $k_{\text{up}} = 2$ E-values	1.4	0.03	16.7
WU-BLAST2 P-values	1.1	0.003	17.5
BLAST P-values	1.0	0.00016	14.8

*Times are from large database searches with genome proteins.

extent of errors. Second, SSEARCH, WU-BLAST2, and FASTA ktup = 1 perform best, though BLAST and FASTA ktup = 2 detect most of the relationships found by the best procedures and are appropriate for rapid initial searches.

The homologous proteins that are found by sequence comparison can be distinguished with high reliability from the huge number of unrelated pairs. However, even the best database searching procedures tested fail to find the large majority of distant evolutionary relationships at an acceptable error rate. Thus, if the procedures assessed here fail to find a reliable match, it does not imply that the sequence is unique; rather, it indicates that any relatives it might have are distant ones.**

**Additional and updated information about this work, including supplementary figures, may be found at <http://sss.stanford.edu/sss/>.

The authors are grateful to Drs. A. G. Murzin, M. Levitt, S. R. Eddy, and G. Mitchison for valuable discussion. S.E.B. was principally supported by a St. John's College (Cambridge, UK) Benefactors' Scholarship and by the American Friends of Cambridge University. S.E.B. dedicates his contribution to the memory of Rabbi Albert T. and Clara S. Bilgray.

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403-410.
- Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* 266, 460-480.
- Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444-2448.
- Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995) *J. Mol. Biol.* 247, 536-540.
- Brenner, S. E., Chothia, C., Hubbard, T. J. P. & Murzin, A. G. (1996) *Methods Enzymol.* 266, 635-643.
- Pearson, W. R. (1991) *Genomics* 11, 635-650.
- Pearson, W. R. (1995) *Protein Sci.* 4, 1145-1160.
- Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* 147, 195-197.
- George, D. G., Hunt, L. T. & Barker, W. C. (1996) *Methods Enzymol.* 266, 41-59.
- Vogt, G., Etzold, T. & Argos, P. (1995) *J. Mol. Biol.* 249, 816-831.
- Henikoff, S. & Henikoff, J. G. (1993) *Proteins* 17, 49-61.
- Bairoch, A. & Apweiler, R. (1996) *Nucleic Acids Res.* 24, 21-25.
- Bairoch, A., Bucher, P. & Hofmann, K. (1996) *Nucleic Acids Res.* 24, 189-196.
- Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* 89, 10915-10919.
- Dayhoff, M., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. (National Bio-medical Research Foundation, Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345-352.
- Brenner, S. E. (1996) Ph.D. thesis. (University of Cambridge, UK).
- Sander, C. & Schneider, R. (1991) *Proteins* 9, 56-68.
- Johnson, M. S. & Overington, J. P. (1993) *J. Mol. Biol.* 233, 716-738.
- Barton, G. J. & Sternberg, M. J. E. (1987) *Protein Eng.* 1, 89-94.
- Lesk, A. M., Levitt, M. & Chothia, C. (1986) *Protein Eng.* 1, 77-78.
- Arratia, R., Gordon, L. & M. W. (1986) *Ann. Stat.* 14, 971-993.
- Karlin, S. & Altschul, S. F. (1990) *Proc. Natl. Acad. Sci. USA* 87, 2264-2268.
- Karlin, S. & Altschul, S. F. (1993) *Proc. Natl. Acad. Sci. USA* 90, 5873-5877.
- Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994) *Nat. Genet.* 6, 119-129.
- Pearson, W. R. (1996) *Methods Enzymol.* 266, 227-258.
- Lipman, D. J., Wilbur, W. J., Smith, T. F. & Waterman, M. S. (1984) *Nucleic Acids Res.* 12, 215-226.
- Wootton, J. C. & Federhen, S. (1996) *Methods Enzymol.* 266, 554-571.
- Waterman, M. S. & Vingron, M. (1994) *Stat. Science* 9, 367-381.
- Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* 13, 669-678.
- Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987) in *Crystallographic Databases: Information Content, Software Systems, Scientific Applications*, eds. Allen, F. H., Bergerhoff, G. & Sievers, R. (Data Comm. Intl. Union Crystallogr., Cambridge, UK), pp. 107-132.
- Brenner, S. E., Chothia, C. & Hubbard, T. J. P. (1997) *Curr. Opin. Struct. Biol.* 7, 369-376.
- Orengo, C., Michie, A., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. (1997) *Structure (London)* 5, 1093-1108.
- Zweig, M. H. & Campbell, G. (1993) *Clin. Chem.* 39, 561-577.
- Gribnikov, M. & Robinson, N. L. (1996) *Comput. Chem.* 20, 25-33.
- Fitch, W. M. (1966) *J. Mol. Biol.* 16, 9-16.
- Chung, S. Y. & Subbiah, S. (1996) *Structure (London)* 4, 1123-1127.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* 25, 3389-3402.
- Girling, R., Schmidt, W., Jr., Houston, T., Amma, E. & Huisman, T. (1979) *J. Mol. Biol.* 131, 417-433.
- Spezio, M., Wilson, D. & Karplus, P. (1993) *Biochemistry* 32, 9906-9916.
- Sayle, R. A. & Milner-White, E. J. (1995) *Trends Biochem. Sci.* 20, 374-376.

Translation**PATENT COOPERATION TREATY****PCT****INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY**
(Chapter II of the Patent Cooperation Treaty)

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 0000054379		FOR FURTHER ACTION		See Form PCT/IPEA/416																								
International application No. PCT/EP2004/002582		International filing date (day/month/year) 12.03.2004	Priority date (day/month/year) 19.03.2003																									
International Patent Classification (IPC) or national classification and IPC																												
Applicant BASF AKTIENGESELLSCHAFT																												
<p>1. This report is the international preliminary examination report, established by this International Preliminary Examining Authority under Article 35 and transmitted to the applicant according to Article 36.</p> <p>2. This REPORT consists of a total of <u>5</u> sheets, including this cover sheet.</p> <p>3. This report is also accompanied by ANNEXES, comprising:</p> <p>a. <input type="checkbox"/> (sent to the applicant and to the International Bureau) a total of _____ sheets, as follows:</p> <p><input type="checkbox"/> sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications authorized by this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions).</p> <p><input type="checkbox"/> sheets which supersede earlier sheets, but which this Authority considers contain an amendment that goes beyond the disclosure in the international application as filed, as indicated in item 4 of Box No. I and the Supplemental Box.</p> <p>b. <input type="checkbox"/> (sent to the International Bureau only) a total of (indicate type and number of electronic carrier(s)) _____, containing a sequence listing and/or tables related thereto, in computer readable form only, as indicated in the Supplemental Box Relating to Sequence Listing (see Section 802 of the Administrative Instructions).</p> <p>4. This report contains indications relating to the following items:</p> <table border="0"><tr><td><input checked="" type="checkbox"/></td><td>Box No. I</td><td>Basis of the report</td></tr><tr><td><input type="checkbox"/></td><td>Box No. II</td><td>Priority</td></tr><tr><td><input type="checkbox"/></td><td>Box No. III</td><td>Non-establishment of opinion with regard to novelty, inventive step and industrial applicability</td></tr><tr><td><input type="checkbox"/></td><td>Box No. IV</td><td>Lack of unity of invention</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Box No. V</td><td>Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability, citations and explanations supporting such statement</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Box No. VI</td><td>Certain documents cited</td></tr><tr><td><input type="checkbox"/></td><td>Box No. VII</td><td>Certain defects in the international application</td></tr><tr><td><input type="checkbox"/></td><td>Box No. VIII</td><td>Certain observations on the international application</td></tr></table>					<input checked="" type="checkbox"/>	Box No. I	Basis of the report	<input type="checkbox"/>	Box No. II	Priority	<input type="checkbox"/>	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability	<input type="checkbox"/>	Box No. IV	Lack of unity of invention	<input checked="" type="checkbox"/>	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability, citations and explanations supporting such statement	<input checked="" type="checkbox"/>	Box No. VI	Certain documents cited	<input type="checkbox"/>	Box No. VII	Certain defects in the international application	<input type="checkbox"/>	Box No. VIII	Certain observations on the international application
<input checked="" type="checkbox"/>	Box No. I	Basis of the report																										
<input type="checkbox"/>	Box No. II	Priority																										
<input type="checkbox"/>	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability																										
<input type="checkbox"/>	Box No. IV	Lack of unity of invention																										
<input checked="" type="checkbox"/>	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability, citations and explanations supporting such statement																										
<input checked="" type="checkbox"/>	Box No. VI	Certain documents cited																										
<input type="checkbox"/>	Box No. VII	Certain defects in the international application																										
<input type="checkbox"/>	Box No. VIII	Certain observations on the international application																										
Date of submission of the demand		Date of completion of this report																										
Name and mailing address of the IPEA/EP		Authorized officer																										
Facsimile No.		Telephone No.																										

Form PCT/IPEA/409 (cover sheet) (January 2004)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/EP2004/002582

Box No. 1

Basis of the report

1. With regard to the language, this report is based on the international application in the language in which it was filed, unless otherwise indicated under this item.
- ☐ This report is based on translations from the original language into the following language _____ which is the language of a translation furnished for the purposes of:
- ☐ international search (Rule 12.3 and 23.1(b))
- ☐ publication of the international application (Rule 12.4)
- ☐ international preliminary examination (Rule 55.2 and/or 55.3)
2. With regard to the elements of the international application, this report is based on (replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report):
- ☐ the international application as originally filed/furnished
- ☒ the description:
- pages 1-44 _____ as originally filed/furnished
- pages* _____ received by this Authority on _____
- pages* _____ received by this Authority on _____
- ☒ the claims:
- nos. 1-30 _____ as originally filed/furnished
- nos.* _____ as amended (together with any statement) under Article 19
- nos.* _____ received by this Authority on _____
- nos.* _____ received by this Authority on _____
- ☒ the drawings:
- sheets 1/2-2/2 _____ as originally filed/furnished
- sheets* _____ received by this Authority on _____
- sheets* _____ received by this Authority on _____
- ☐ a sequence listing and/or any related table(s) - see Supplemental Box Relating to Sequence Listing.
3. ☐ The amendments have resulted in the cancellation of:
- ☐ the description, pages _____
- ☐ the claims, nos. _____
- ☐ the drawings, sheets/figs _____
- ☐ the sequence listing (specify): _____
- ☐ any table(s) related to sequence listing (specify): _____
4. ☐ This report has been established as if (some of) the amendments annexed to this report and listed below had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).
- ☐ the description, pages _____
- ☐ the claims, nos. _____
- ☐ the drawings, sheets/figs _____
- ☐ the sequence listing (specify): _____
- ☐ any table(s) related to sequence listing (specify): _____

* If item 4 applies, some or all of those sheets may be marked "superseded."

Form PCT/PEPA/409 (Box No. 1) (January 2004)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/EP2004/002582

Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Claims 1-30	YES
	Claims	NO
Inventive step (IS)	Claims	YES
	Claims 1-30	NO
Industrial applicability (IA)	Claims 1-30	YES
	Claims	NO

2. Citations and explanations (Rule 70.7)

1. Reference to the following documents:

D1: EP-A-O 486 290 (AMOCO CORP)

20 May 1992 (1992-05-20)

D2: DE 197 44 212 A (SCHERING AG)

15 April 1999 (1999-04-15).

2. The present application fails to satisfy the requirements of PCT Article 33(2) because the subject matter of claims 1-30 lacks novelty (PCT Article 33(2)).

3.1 Document D1 discloses (the references in brackets are to said document) a method for the production of a variety of sterols which can be regarded as intermediate and/or resultant products of ergosta-5,7-dienol, said method involving the cultivation of organisms which have a reduced Δ^{22} -desaturase (erg5) activity and an increased HMG-CoA-reductase activity relative to the wild type (see claim 13).

3.2 Document D2 discloses a method for the production of a variety of sterols which can be regarded as

/...

Form PCT/IPEA/409 (Box No. V) (January 2004)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/EP2004/002582

Box No. V

Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability, citations and explanations supporting such statement

intermediate and/or resultant products of ergosta-5,7-dienol, said method involving the cultivation of organisms which have an increased HMG-CoA-reductase activity and increased squalene poxidase (erg1) activity relative to the wild type (see claim 2.a-iv)).

- 4.1 The present application fails to satisfy the requirements of PCT Article 33(1) because the subject matter of claims 1-30 does not involve an inventive step (PCT Article 33(3)).
- 4.2 Document D1 is considered to be the prior art closest to the subject matter of claim 1 and the subject matter of claim 1 therefore differs from D1 in that the method is implemented using organisms which, by comparison with the present claims, additionally display an increased activity of a further enzyme, which enzyme can be selected from the group comprising lanosterol-C14-demethylase (erg11), squalene poxidase (erg1) and squalene synthetase (erg9).
- 4.3 The problem addressed by the present invention can consequently be regarded as that of providing an alternative method for producing ergosta-5-7-dienol (and/or biosynthetic intermediate and/or resultant products thereof).

/...

Form PCT/FEA/409 (Box No. V) (January 2004)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/EP2004/002582

Box No. V

Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability, citations and explanations supporting such statement

- 4.4 The solution proposed in claims 1-10 and 15-18 of the present application cannot be considered inventive (PCT Article 33(3)) for the following reasons:

as stated in point 3.2 above, D2 discloses (see claim 2.a-iv)) a method for the production of ergosta-5-7-dienol and/or biosynthetic intermediate and/or resultant products thereof by means of an increased t-HMG and erg1 activity. A person skilled in the art could therefore arrive at the solution to the present problem by combining the disclosures of D1 and D2, without unreasonable experimental input.

- 4.5 Moreover, the problem cannot be considered to have been solved for the entire scope of protection claimed in claim 1: (i) it has not been shown that the problem has been solved for all intermediate and/or resultant products of ergosta-5,7-dienol. Tables 2 and 3 (data for *S. cerevisiae* GRFtHIura3ERG1erg5) show the decrease in the content of squalene (which can be considered to be an intermediate product of ergosta-5-7-dienol) by comparison with table 1 (data for *S. cerevisiae* GRFtHIura3) and table 3 (data for *S. cerevisiae* GRFtHIura3erg5); (ii) nor has it been shown that the aforementioned problem has been solved by an increase in lanosterol-C14-demethylase (erg11) or squalene synthetase (erg9) activity (in addition

/...

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/EP2004/002582

Box No. V

Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability;
citations and explanations supporting such statement

to a reduction in erg5 activity and an increase in
HMG reductase activity).

- 4.6 In consequence, claims 1-30 do not involve an
inventive step (PCT Article 33(3)).

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/EP2004/002582

Box No. VI

Certain documents cited

1. Certain published documents (Rule 70.10)

Application No.
Patent No.Publication date
(day/month/year)Filing date
(day/month/year)Priority date (valid claim)
(day/month/year)

2. Non-written disclosures (Rule 70.9)

Kind of non-written disclosure

Date of non-written disclosure
(day/month/year)Date of written disclosure
referring to non-written disclosure
(day/month/year)

Form PCT/IPEA/409 (Box No. VI) (January 2004)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/EP2004/002582

Supplemental Box

In case the space in any of the preceding boxes is not sufficient,
Continuation of:

Box VI

Document D4 (WO03/064650), cited in the search report,
might be prejudicial to novelty in a later regional
phase.

Form PCT/IPEA/409 (Supplemental Box) (January 2004)

U.S. Serial No.: 10/549,871
Examiner: Mohammad Y. Meah
GAU: 1652
Page 14 of 14

RECEIVED
CENTRAL FAX CENTER
AUG 13 2007

upon this difference, Weber does not anticipate the present application. The Applicant respectfully requests that this rejection be withdrawn.

CONCLUSION

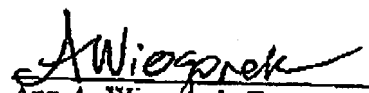
In light of the foregoing Amendments and remarks, it is believed that the rejections and objections of record have been obviated, and allowance of this application is respectfully solicited. If a telephone conference would facilitate examination of this application in any way, the examiner is invited to contact the applicant's attorney at (619) 846-4850. The Examiner's consideration of this matter is gratefully acknowledged.

FEES

The Commissioner is authorized to charge the fees for a petition for a three-month extension of time for a large entity (\$1020) and any other fees deemed necessary in connection with the above-application to Deposit Account No. 50-1047.

Respectfully submitted,

Attorney for Applicant
Mayer & Williams PC
251 North Avenue West, 2nd Floor
Westfield, NJ 07090
Tel.: 619-846-4850
Fax: (908) 518-7795


Ann A. Wieczorek, Esq.
Registration No. 46,087

<p align="center">Certificate of Facsimile Transmission</p> <p>I hereby certify that this correspondence and any document referenced herein are being sent to the United States Patent and Trademark office via Facsimile to: 571-273-8300 on <u>8/13/07</u></p> <p align="center"><u>Marjorie Scariati</u> (Printed Name of Person Sending Correspondence)</p> <p align="center"><u>Marjorie Scariati</u> (Signature)</p>
